

⑩日本国特許庁 (JP)

⑪特許出願公開

⑫公開特許公報 (A) 平4-45781

⑬Int.Cl.⁵

C 12 M 1/00
C 07 K 3/00

識別記号

庁内整理番号

A 8717-4B
7731-4H

⑭公開 平成4年(1992)2月14日

審査請求 未請求 請求項の数 1 (全7頁)

⑮発明の名称 タンパク質構造比較装置

⑯特 願 平2-154397

⑰出 願 平2(1990)6月13日

⑱発明者 長谷川 淳 神奈川県川崎市中原区上小田中1015番地 富士通株式会社
内

⑲出願人 富士通株式会社 神奈川県川崎市中原区上小田中1015番地

⑳代理人 弁理士 井桁 貞一 外2名

明細書

1. 発明の名称

タンパク質構造比較装置

2. 特許請求の範囲

プライマ側およびターゲット側の各タンパク質の一次構造データを保持する保持手段と、

前記一次構造データを比較して類似するアミノ酸配列を抽出する抽出手段と、

該抽出結果に基づいて前記ふたつのタンパク質の立体構造表示モデルを重ね合わせる際の基準となる残基位置を指定する指定手段と、を備えることを特徴とするタンパク質構造比較装置。

3. 発明の詳細な説明

〔概要〕

タンパク質構造比較装置に関し、

一次構造を参照することにより、基準原子の指

定を簡素化することを目的とし、

プライマ側およびターゲット側の各タンパク質の一次構造データを保持する保持手段と、前記一次構造データを比較して類似するアミノ酸配列を抽出する抽出手段と、該抽出結果に基づいて前記ふたつのタンパク質の立体構造表示モデルを重ね合わせる際の基準となる残基位置を指定する指定手段と、を備える。

〔産業上の利用分野〕

本発明は、タンパク質構造比較装置、特にタンパク質の立体構造を解析する装置であって、コンピュータグラフィックス技術を応用したタンパク質構造比較装置に関する。

生体における生物的機能は、主にタンパク質とその結合体によって担われている。タンパク質のなかでも生物に固有の働きをもつものの代表は酵素であるが、そのほかにも、抗原抗体反応に関与するイムノグロブリンや、呼吸に関与するヘモグロビン、生体運動に関与するミオシンやアクチン、

それに、ホルモン・タンパク質など、多種多様なタンパク質が知られている。

すなわち、生体の各生理機能はその系に固有のタンパク質によって担われ、また、タンパク質の働きは、結合組織の構成物であるコラーゲンや、毛や爪の成分であるケラチンなど硬タンパク質と総称されるもののように生体構造の維持・保護に役立っているものがあり、さらに、筋肉タンパクのように生理的機能だけでなく構造形成に関与しているものもある。

こうしたタンパク質固有の機能は、それぞれに固有の立体構造によって保証され、かつ、他の物質との相互作用の結果生じる特異的な構造変化によってその機能が発現される。

したがって、任意のタンパク質を解析する作業は、その立体構造を知ることに他ならない。

【従来の技術】

従来のタンパク質構造比較装置としては、例えば、X線結晶解析技術とコンピュータグラフィッ

像拡大や回転といった操作を繰り返す必要があり、作業が繁雑になるといった問題点があった。

すなわち、第7図に示すように簡単な構造（分子量が少ない）のタンパク質であれば、基準原子（図中○印）の指定は容易であり、第8図に示すように最小二乗法による重ね合わせ結果も良好に得られるが、例えば、第9図に示すように複雑な構造のタンパク質（多くのタンパク質はその構造が複雑）の場合には、相当の試行錯誤を繰り返さなければならぬ。

本発明は、このような問題点に鑑みてなされたもので、一次構造を参照することにより、基準原子の指定を簡素化することを目的としている。

【課題を解決するための手段】

本発明は、上記目的を達成するためその原理構成図を第1図に示すように、プライマ側およびターゲット側の各タンパク質の一次構造データを保持する保持手段と、前記一次構造データを比較して類似するアミノ酸配列を抽出する抽出手段と、

クス技術とを応用した装置が知られている。

この装置では、解析しようとするタンパク質（以下、ターゲット）の立体構造（一般に3次元構造）と既知のタンパク質（以下、プライマ）の立体構造とをグラフィックス上で重ね合わせ、構造上の微細な違いからターゲットの機能を解析する。正確な重ね合わせのためには、コンピュータによる数値解析、例えば、間数の二乗和を最小にする変数の値を求める最小二乗法（least squares method）を実行する必要があり、かかる最小二乗法を行うには、予めターゲットおよびプライマの主鎖を構成する多數の原子のなかから、少なくとも4つの原子（重ね合わせの基準となる原子、以下、基準原子）を指定しておく必要がある。

【発明が解決しようとする課題】

しかしながら、上記従来のタンパク質構造比較装置にあっては、基準原子の指定をオペレータの直感にたよって発見的に行うものであったため、複雑な立体構造を持つものに対しては部分的な画

像抽出結果に基づいて前記ふたつのタンパク質の立体構造表示モデルを重ね合わせる際の基準となる残基位置を指定する指定手段と、を備える。

【作用】

タンパク質の一次構造とは、多數のアミノ酸をペプチド結合でつなげたポリペプチド鎖の何本かの集合において、そのアミノ酸の配列順序をいう。一次構造の一端にはアミノ基が、他端にはカルボキシル基が存在するので前者をN末端、後者をC末端という。N末端からC末端までつながる各アミノ酸に、その種類を表すアルファベット符号と番号（1～n）が付与されている。1955年にF. Sangerがインシュリンについての一次構造を明らかにしたのを始め現在までに数千種類のタンパク質の一次構造が判明している。

すなわち、一次構造は、タンパク質に固有のものであるから、一次構造を参照することにより、基準原子の指定を簡単化することができる。

(実施例)

以下、本発明を図面に基づいて説明する。

第2～6図は本発明に係るタンパク質構造比較装置の一実施例を示す図である。

第2図において、タンパク質構造比較装置10は、外部記憶装置20（保持手段）、コンピュータ30（抽出手段、指定手段）、メモリ40、ディスプレイ50、キーボード60、およびマウス等の座標入力装置70を備える。

外部記憶装置20には、各種タンパク質ごとの固有データ（立体構造データ（原子座標データ）やアミノ酸配列データ（一次構造データ））が格納されており、コンピュータ30による制御の基で、指定された2つのタンパク質に関するデータを取り出し、重ね合わせ処理を実行した後、その結果をディスプレイ50上にグラフィック表示する。

ここで、メモリ40には、所定のプログラムが格納されており、このプログラムは、タンパク質の取り出し指定、指定タンパク質の立体構造データおよび一次構造データの表示、一次構造データ中

の類似アミノ酸配列の指定、類似アミノ酸配列中の残基位置指定、最小二乗法による重ね合わせ演算などを実行する処理プログラムである。なお、タンパク質や残基位置の指定は、キーボード60および座標入力装置70の操作により行う。

第3図は本実施例の全体の処理を示すフローチャートである。

このフローチャートにおいて、ステップS1で任意の2つのタンパク質（プライマ側およびターゲット側タンパク質）を指定すると、ステップS2で、指定した2つのタンパク質の立体構造がディスプレイ50上に表示される。ここでは、第9図をその表示例とする。図中の左半分がプライマ側タンパク質の立体構造（3次元構造）、右半分がターゲット側タンパク質のそれである。

次いで、ステップS3では、ターゲット側およびプライマ側タンパク質の一次構造を表示する。表示例は第4図に示される。図中上半分がターゲット側タンパク質の一次構造、下半分がプライマ側タンパク質のそれである。

一次構造（表）は、アミノ酸のタイプを表すアルファベット符号が配列順に左から右へ上から下へと並べられ、かつ、配列順に番号（但し、10の単位のみ表記）が付されている。例えば、プライマ側の番号11は符号Yで表わされるアミノ酸、ターゲット側の番号81は符号Aで表されるアミノ酸である。

ところで、2つの一次構造を仔細に比較すると、いくつかの部分で類似性が見られる。この例では、第5図（a）に示す抽出部分について類似している。すなわち、

①ターゲット側の番号72から番号85までのアミノ酸配列【T W S I S Y G D G S S A S G】がプライマ側の番号70から番号83までのアミノ酸配列と一致し、

②ターゲット側の番号119から番号125までのアミノ酸配列【N D G L L G L】がプライマ側の番号118から番号124までのアミノ酸配列と一致し、

③ターゲット側の番号181から番号184までの

アミノ酸配列【C S L T】がプライマ側の番号177から番号180までのアミノ酸配列と一致し、

④ターゲット側の番号218から番号224までのアミノ酸配列【D T G T T L L】がプライマ側の番号213から番号219までのアミノ酸配列と一致し、

⑤ターゲット側の番号308から番号311までのアミノ酸配列【Y V V F】がプライマ側の番号307から番号310までのアミノ酸配列と一致している。

ステップS4では、①～⑤の各アミノ酸配列ごとにひとつのアミノ酸（残基）を選択してその番号を指定する。すなわち、残基位置を指定する。例えば、①にあっては符号Iを選択してその番号75／73（ターゲット側／プライマ側、以下同様）を指定し、②にあっては符号Lを選択してその番号123／122を指定し、③にあっては符号Sを選択してその番号183／179を指定し、④にあっては符号Gを選択してその番号224／219を指定し、⑤にあっては符号Vを選択してその番号310／309を指定する。これらの残基位置を2つのタンパク

質の立体構造上に示すと、第5図(b)のようになる。

最後に、ステップS5で最小二乗法による重ね合わせ演算を実行した後、ステップS6でその結果を表示し、処理を終了する。

重ね合わせ結果は、例えば第6図のようになり、この画像を検証して構造上の微妙な相違を見つけることで、ターゲット側タンパク質の構造を解析する。

以上のように、本実施例によれば、タンパク質の構造を考える上で欠かすことのできないアミノ酸配列(一次構造)を参照し、重ね合わせの基準となる残基(基準原子)位置を指定するようにしたので、従来のように発見的かつ直感的な指定方法に比べて指定作業を大幅に簡素化することができる。

〔発明の効果〕

本発明によれば、タンパク質の一次構造を参照して基準原子を指定したので、指定を簡素化する

ことができる。

4. 図面の簡単な説明

第1図は本発明の概念構成図、

第2~6図は本発明に係るタンパク質構造比較装置の一実施例を示す図であり、

第2図はそのシステム構成図、

第3図はその動作フローチャート、

第4図はその2つのタンパク質の一次構造を示す図、

第5図(a)はその一次構造の類似する部分を抜粋して示す図、

第5図(b)はその2つのタンパク質の立体構造と指定残基位置を示す図、

第6図はその重ね合わせ結果を示す図である。

第7~9図は従来例を示す図であり、

第7図はその簡単な構造の2つのタンパク質を示す図、

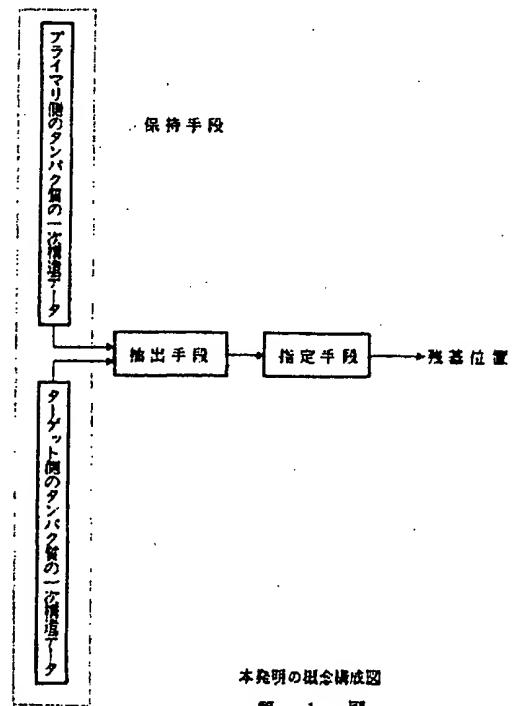
第8図はその重ね合わせ結果を示す図、

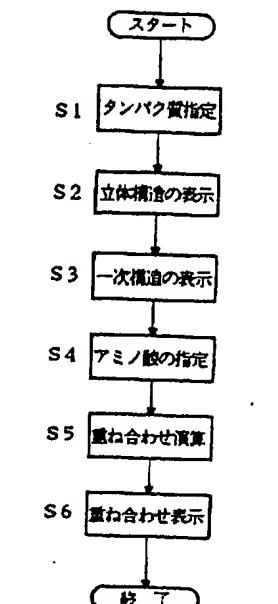
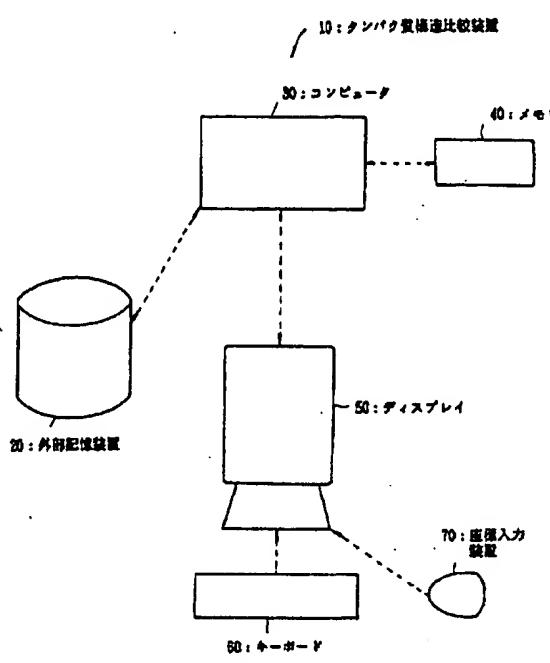
第9図はその複雑な構造の2つのタンパク質を

示す図である。

20……外部記憶装置(保持手段)、
30……コンピュータ(抽出手段、指定手段)。

代理人弁理士 井桁貞





(ターゲット側)

```

+ 10 + 20 + 30 + 40 + 50
AGITGTVPMF YKNDIPEYCG VTIIGTPGKHF ALDPTDTESSD LHLIASTLCYH
+ 60 + 70 + 80 + 90 + 100
CGSGCTKTYDP KNGSTTQADG RTWISIYGDG SSASCILAKD KYNALGGLIK
+ 110 + 120 + 130 + 140 + 150
GATIGELAKRE AASFASPCPQD GLLCLCPDTI TTVEGKTPH DNLSISCLIS
+ 160 + 170 + 180 + 190 + 200
RPIPCVVLCK AKNGGCGEYI PGGTGTDTPR GSLITWPIDM SECWAGCTIVD
+ 210 + 220 + 230 + 240 + 250
RAYTGISTVVA SSPDGILDTG TILLIIPKNT AASVARAYGA SONGDGTYYI
+ 260 + 270 + 280 + 290 + 300
SCDTISAFKPL VPSTINACSPF VPSDLSVPEE FGCGCIACPG KGMGFAIG
+ 310 + 320
DPLIENNYVV PMAQGVPVYD! APTAE

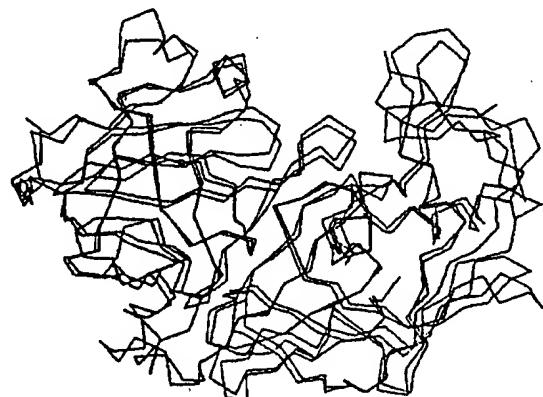
```

(プライマリ)

```

+ 10 + 20 + 30 + 40 + 50
AASCGATMTP TANDEEYIP VTIIGTPGKHF ALDPTDTESSD VFTSTELPASD
+ 60 + 70 + 80 + 90 + 100
GSGCTKTYDP ATCKELSCYI WSISIGCGSS ASCGVPTVSIY TVCGVTAEGG
+ 110 + 120 + 130 + 140 + 150
AVDAGDQFSA QPGQDTMDCG LLGLAPSSIN TTWPLSOTTF PDTVKSLSLQ
+ 160 + 170 + 180 + 190 + 200
PLPVALKHA QPGVYDPEFI DSSKTVTSLI YTCGIVMSLDF NSPMDVSTTA
+ 210 + 220 + 230 + 240 + 250
GSLSGCGPSC IADTGTTLIL LDGSVYVSGTY SGSVCGAODS HAGGYVFDG
+ 260 + 270 + 280 + 290 + 300
TKLPDPVSYI SGTATVPGS LIVYCPGSCG STCGAGLQSY SGCGPSIFG
+ 310 + 320
IPLESQYVVP DSGCPDGLCPA PDA

```

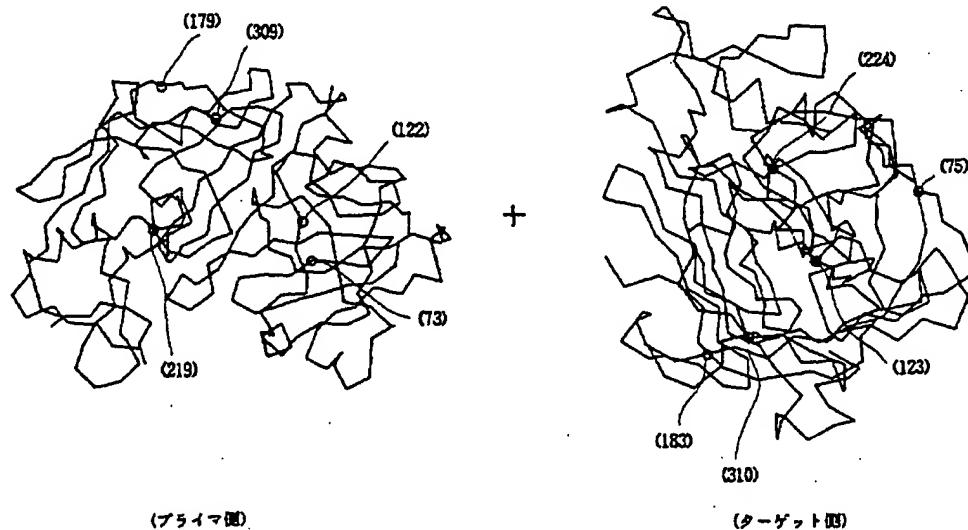


実施例の2つのタンパク質の一次構造を示す図
第4図

②		①		②		①		②		①	
+	10	+	20	+	30	+	40	+	50		
+	60	+	70	+	80	+	90	+	100		
+	110	+	120	+	130	+	140	+	150		
+	160	+	170	+	180	+	190	+	200		
+	210	+	220	+	230	+	240	+	250		
+	260	+	270	+	280	+	290	+	300		
+	310	+	320	+							

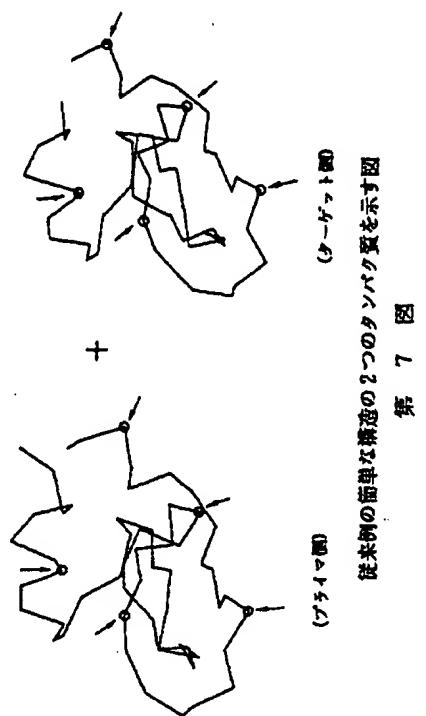
(ターゲット側)

(プライマ側)

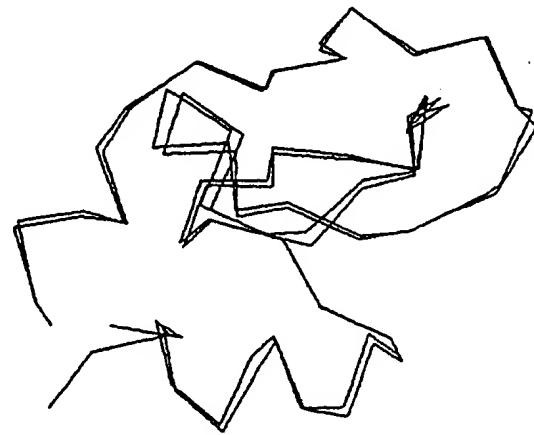


—実施例の2つのタンパク質の立体構造と指定残基位置を示す図

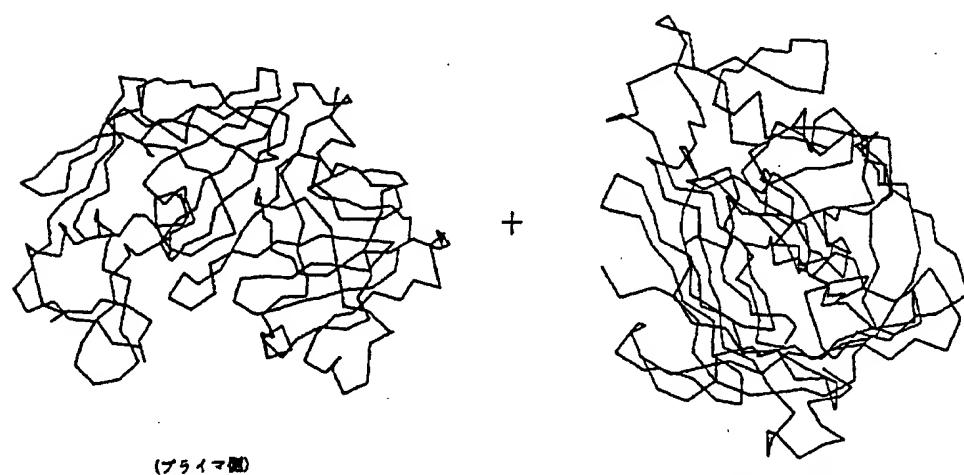
第5図 (b)



従来例の簡単な構造の2つのタンパク質を示す図
第7図



従来例の重ね合わせ結果を示す図
第8図



従来例の複雑な構造の2つのタンパク質を示す図
第9図